

# 大量ボーリングデータの地図展開と高速検索機能の開発

北尾馨\*・中田文雄\*\*・土屋彰義\*\*・池田俊雄\*\*

## Development of plotting-system and search-engine for mass boring data on web-based mapping system

Kaoru Kitao\*, Fumio Nakada\*\*, Akiyoshi Tsuchiya\*\* and Toshio Ikeda\*\*

\* 合資会社キューブワークス CubeWorks Inc., 4-1, Mizukino-2chome, Moriya, Ibaraki 302-0121, Japan. E-mail: kitao@cubeworks.co.jp

\*\* 一般社団法人全国地質調査業協会連合会 Japan Geotechnical Consultants Association, 1-5-13, Uchikanda, Chiyoda-ku, Tokyo, 101-0047, Japan.

キーワード：全文検索, PostgreSQL, PGroonga, パーティショニング, 関数インデックス

Key words : Full Text Search, PostgreSQL, PGroonga, Partitioning, Indexes on Expressions

### 1. はじめに

本件は、情報流通連携基盤の地盤情報における実証事務局（以下、実証事務局という）が運営する「全国ボーリング所在情報公開サイト」において、収集した全てのボーリングデータをウェブ地図上で高速に展開、検索可能なシステム（以下、本件システムという）を開発したので、これを紹介するものである。

本件システムの開発に先立ち、著者らは「こうち地盤情報公開サイト」として高知県のボーリングデータ約 6,700 件をウェブ地図上で展開、検索可能なシステム（以下、こうちシステムという）を 2015 年に開発したが、このシステムではデータベース (PostgreSQL) が標準で装備する機能のみを利用して、高速に動作するシステムを構築することができた。本件システム開発の検証作業として、こうちシステムと同等のシステムを用意し、実際に使用する約 20 万件のボーリングデータを使って動作確認を実施したが、データ提供元や掘削深度などの条件指定による検索は十分高速に動作したものの、発注者や地質名、住所等に対する文字列検索機能を使用した場合に応答速度の著しい劣化が認められた。この問題を解決するために構築したシステムが本件の成果である。

### 2. 手法

本件システムは北尾・西岡 (2014) の手法を採用している。その手法とは、ボーリングデータの位置情報 (経緯度) とは別に、地図上にその位置を高速展開するためのデータ (以下、高速展開用データという) を別途保持 (データを格納するテーブルを、高速展開用テーブルという) し、地図が描画される都度データベースから該当データを検索して地図上にその散布状況を展開する仕組みである。この仕組みに、下記にて説明する機能を組み合わせて使用した。

#### 2.1 全文検索エンジン PGroonga

PGroonga は全文検索エンジン groonga の仕組みを使って PostgreSQL のインデックスを作成するためのプログラムで、ページラングと読む。

一般に、PostgreSQL や MySQL といったリレーショナルデータベースで文字列検索を実行する場合、検索文字列に対する対象文字列の完全一致もしくは前方一致であれば、データベース標準のインデックス機能を使って高速な検索を実行することが可能である。しかし中間一致や後方一致で検索する場合、標準のインデックス機能のみを用いて高速な検索を実行することができず、これらを実現するためには別途追加でプログラムを組み込む必要がある (英文等単語間の区切りが明瞭である場合はこの限りでない)。

本件システムの開発において、高速な全文検索機能を実現するためのプログラムとして、pg\_trgm, pg\_bigm, PGroonga の各プログラムについてそれぞれ検証した。

pg\_trgm (ピージートリグラム) は PostgreSQL に付属するプログラムであるが、デフォルトではインストールされず、日本語環境でこれを使用する場合、PostgreSQL のソースコードを変更してコンパイル、インストールする必要がある。また pg\_trgm はその仕様上、検索文字列が 2 文字以下の場合、基本的には全文検索インデックスを使用することができず高速な動作を見込めない。よって本件システムではこれを不採用とした。

pg\_bigm (ピージーバイグラム) は株式会社 NTT データが開発した PostgreSQL ライセンスで配布されるプログラムで、pg\_trgm をベースに開発された日本語対応の全文検索プログラムである。pg\_bigm は 2 文字以下の文字列で検索した場合でも全文検索用インデックスを使用することができ、pg\_trgm と比較して日本語との親和性がより高いといえる。本件では、検証段階において pg\_bigm をインストールしてプログラムを調整、試用したが、検索結果の該当件数が多い場合、良好な応答速度を得ることができず、結果として pg\_bigm の採用を見送った。

対して PGroonga は PGroonga Project が開発する、PostgreSQL に転置インデックス機能を提供するプログラムである (PostgreSQL ライセンス)。転置インデックスとは、書籍等の索引をコンピュータ上で実現するための技術である。PGroonga を本件システムに組み込んで試用した

結果、良好な応答速度で検索可能であったため、本件では PGroonga を全文検索エンジンとして採用した。

## 2.2 データベース設計と関数インデックス

本件システムでは、データ提供元や調査件名等複数の項目（カラム）にわたって文字列で検索する機能を提供する必要がある。前述の PGroonga を用いて各項目にそれぞれ全文検索用インデックスを作成することも可能であるが、プログラムが煩雑になる。本件では複数の項目を一つに束ねて検索することを可能にする関数インデックスを採用した。これは予め検索したい項目を一つに束ねて返す関数を用意し、この関数の返り値に対して全文検索用インデックスを作成する手法である。

こうしシステムでは、データベースを正規化する目的で、ボーリングデータを保持するテーブルとは別に複数のサブテーブルを用意して外部キーで結合し、ボーリングデータの一部をサブテーブルに格納する手法を採用していた。本件システムではサブテーブルを廃止し、PostgreSQL 独特のデータ型である配列型のカラムを使ってデータを格納することとした。これにより検索時に複数のテーブルを結合する必要がなく、結果として応答速度の向上を見込むことができる。本件システムで作成した配列型カラムの一部は文字列検索の対象とする必要があり、全文検索用インデックスを作成する点においても都合の良い結果となった。

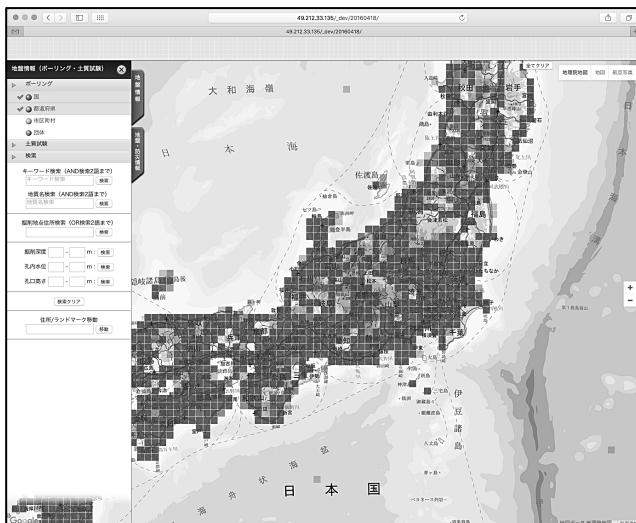
## 2.3 パーティショニング

開発するシステムの仕様に依存するが、北尾・西岡 (2014) の手法をそのまま採用して本件システムを開発する場合、ボーリングデータの件数を 20 万件と仮定すると、高速展開用テーブルには 200 万件のデータが投入される。データベースは通常、テーブルに格納されるデータ量が多くなるとデータ抽出にかかる時間が長くなる傾向がある。本件ではパーティショニングと呼ばれるテーブルを分割する手法を採用し、高速展開用テーブルを 10 のテーブルに分割してデータを分散格納することで、1 つのテーブルで保持するデータの件数が大きくなりすぎないように配慮している。

## 3. 画面設計

本件システムのビューア画面を図 1 に示す。

本件システムを利用するクライアントの要件は一般的なウェブブラウザであることのみで、特別なソフトウェアやウェブブラウザのプラグイン等を別途インストールする必要はない。



第 1 図 本件システムのビューア画面。

操作メニューは画面左に格納し、タブをクリックすることで操作メニューを開閉することができる。

ボーリングデータは、データ提供元を選択することで表示される。データは、低ズームレベル時にはメッシュの濃淡でその分布密度を表現し、高ズームレベル時にはマーカーで個々のボーリング位置を表すよう開発している。

全文検索エンジン PGroonga を用いた文字列検索機能は 3 種類提供し、それらは発注者や受注者、調査件名等に対する 2 語以内でのキーワード検索 (AND 検索)、各ボーリングに含まれる岩相に対する 2 語以内での地質名検索 (AND 検索)、掘削地点の住所に対する 2 語以内での掘削地点住所検索 (OR 検索) である。

その他に、掘削深度、孔内水位、孔口高さの上限、下限を指定して検索する機能を装備している。これらの検索には PostgreSQL に標準で装備されている B-tree インデックスを使用しており、良好な応答速度で動作している。

いずれの検索においてもその応答速度はおおよそ 3 秒以内であり、十分実用的な速度である。

## 4. データの管理

本件システムはデータ管理機能 (管理画面) を備えており、ビューア画面と同じくウェブブラウザから操作する。本件システムで表示するデータは 20 万件以上を想定しており、今後もデータ数が増加すると考えられる。データの追加登録において、地方自治体等団体単位での一括登録が想定されるが、本件システムでは、予め定められた書式のエクセルテンプレートでデータの一覧を作成し、これをエクスポートして管理画面からサーバにアップロードして登録する機能を装備している。これにより 1 度の操作で最大 5,000 件のデータを登録することが可能である。

## 5. おわりに

本件システムは、情報流通連携基盤の地盤情報における実証に係わる請負 (総務省情報流通行政局情報流通振興課 2012 年発注) で収集整理された地盤情報を利用しており、実証事務局が運営するウェブサイト「全国ボーリング所在情報公開サイト」での公開を予定している。

## 文 献

- 北尾馨・西岡芳晴 (2014) FOSS で実現する大量位置データの高速描画処理. 情報地質, vol.25, no.2, pp.110-111.
- 高知地盤情報利用連絡会 (2015) こうち地盤情報公開サイト, <http://geonews.zenchiren.or.jp/kochi/>, 最終アクセス日 2016.5.8
- 情報流通連携基盤の地盤情報における実証事務局 (2015) 全国ボーリングデータ所在情報公開サイト, <http://geonews.zenchiren.or.jp/zenkoku/documents/Zjissyo.html>, 最終アクセス日 2016.5.8
- NTT Data Corporation (2016) pg\_bigm (ピージーバイグラム), <http://pgbigm.osdn.jp/>, 最終アクセス日 2016.5.8
- PGroonga Project (2016) PGroonga – Make PostgreSQL fast full text search platform for all languages!, <https://PGroonga.github.io/>, 最終アクセス日 2016.5.8
- PostgreSQL グローバル開発グループ (2016) PostgreSQL 9.5.2 文書, <https://www.postgresql.jp/document/9.5/html/>, 最終アクセス日 2016.5.8